Elevated amygdala responses during *de-novo* Pavlovian conditioning in alcohol-use disorder are associated with Pavlovian-to-Instrumental transfer and relapse latency

Claudia Ebrahimi, Maria Garbusow, Miriam Sebold, Ke Chen, Michael N. Smolka, Quentin J.M. Huys, Ulrich S. Zimmermann, Florian Schlagenhauf, Andreas Heinz

PII: S2667-1743(23)00011-3

DOI: https://doi.org/10.1016/j.bpsgos.2023.02.003

Reference: BPSGOS 202

To appear in: *Biological Psychiatry Global Open Science*

Received Date: 21 October 2022

Revised Date: 3 February 2023

Accepted Date: 3 February 2023

Please cite this article as: Ebrahimi C., Garbusow M., Sebold M., Chen K., Smolka M.N., Huys Q.J.M., Zimmermann U.S., Schlagenhauf F. & Heinz A., Elevated amygdala responses during *de-novo* Pavlovian conditioning in alcohol-use disorder are associated with Pavlovian-to-Instrumental transfer and relapse latency, *Biological Psychiatry Global Open Science* (2023), doi: https://doi.org/10.1016/j.bpsgos.2023.02.003.

Elevated amygdala responses during *de-novo* Pavlovian conditioning in alcohol-use disorder are associated with Pavlovian-to-Instrumental transfer and relapse latency

Running title: Pavlovian conditioning in AUD

Claudia Ebrahimi[1,2], Maria Garbusow[1], Miriam Sebold[1,3], Ke Chen[1], Michael N. Smolka[2,4], Quentin J. M. Huys[5,6,7], Ulrich S. Zimmermann[2,8], Florian Schlagenhauf*[1], Andreas Heinz*[1,9]

[1] Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Psychiatry and Neurosciences | CCM, NeuroCure Clinical Research Center, Berlin, Germany.
[2] Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Dresden, Germany.
[3] Technische Hochschule Aschaffenburg, University of Applied Sciences, Aschaffenburg, Germany.
[4] Neuroimaging Center, Technische Universität Dresden, Dresden, Germany.
[5] Applied Computational Psychiatry Lab, Mental Health Neuroscience Department, Division of Psychiatry, University College London.
[6] Applied Computational Psychiatry Lab, Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Queen Square Institute of Neurology, University College London.
[7] Camden and Islington NHS Foundation Trust.
[8] Department of Addiction Medicine and Psychotherapy, kbo Isar-Amper Klinikum Region München, Haar, Germany
[9] Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, NeuroCure Cluster of Excellence, Berlin, Germany.

*Contributed equally

Keywords: Pavlovian conditioning, alcohol use disorder, Pavlovian-instrumental transfer, relapse, amygdala, neuroimaging

Corresponding author: Claudia Ebrahimi, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Psychiatry and Neurosciences | CCM, Charitéplatz 1, 10117 Berlin, Germany

Background: Contemporary learning theories of drug addiction ascribe a key role to Pavlovian learning mechanisms in the development, maintenance and relapse of addiction. In fact, cue-reactivity research has demonstrated the power of alcohol-associated cues to activate the brain's reward system, which has been linked to craving and subsequent relapse. However, whether de-novo Pavlovian conditioning is altered in alcohol use disorder (AUD) has been rarely investigated.

Methods: To characterize *de-novo* Pavlovian conditioning in AUD, n=62 detoxified patients with AUD and n=63 matched healthy controls completed a Pavlovian learning task as part of a Pavlovian-to-instrumental transfer (PIT) paradigm during an fMRI session. Patients were followed up for 12-months to assess drinking behavior and relapse status.

Results: While patients and controls did not differ in their ability to explicitly acquire the contingencies between conditioned and unconditioned stimuli, patients with AUD displayed significantly stronger amygdala responses towards Pavlovian cues; an effect primarily driven by stronger BOLD differentiation during learning from reward compared to punishment. Moreover, in patients compared to controls, differential amygdala responses during conditioning were positively related to the ability of Pavlovian stimuli to influence ongoing instrumental choice behavior, measured in a subsequent PIT test. Finally, patients who relapsed within the 12-month follow-up period showed an inverse association between amygdala activity during conditioning and relapse latency.

Conclusions: We provide evidence of altered neural correlates of *de-novo* Pavlovian conditioning in patients with AUD, especially for appetitive stimuli. Thus, heightened processing of Pavlovian cues might constitute a behaviorally relevant mechanism in alcohol addiction.

1. Introduction

Alcohol use disorder (AUD) has been conceptualized as a disorder of maladaptive learning and memory (1–4). The incentive sensitization theory (1,5) highlights the motivational power of environmental stimuli to promote craving, drive recurrent drug use and ultimately increase relapse risk. However, the underlying Pavlovian learning process, whereby initially neutral stimuli (conditioned stimuli, CS+) acquire motivational properties through repeated pairings with the hedonic effects of a reinforcer like alcohol (unconditioned stimulus, US) has been rarely investigated in AUD (6).

Human neuroimaging research elucidated an extended network subserving Pavlovian threat and appetitive conditioning, including the amygdala, hippocampus, ventral striatum (VS) entailing the nucleus accumbens (NAcc), dorsal anterior cingulum (dACC) and orbitofrontal cortex (7–10). In AUD, surprisingly little is known about the underlying Pavlovian learning process, and we are unaware of any imaging studies investigating *de-novo* Pavlovian conditioning with drug- or non-drug rewards in this psychiatric condition. This might be partly due to methodological challenges human appetitive conditioning research is facing (11). In contrast, two fMRI studies used a threat conditioning protocol in AUD patients, providing first evidence for attenuated BOLD responses towards threat-predicting cues: Yang and colleagues (12) found attenuated neural differentiation in pregenual ACC, medial prefrontal cortex (PFC) and posterior cingulate cortex (PCC) between a CS predicting a high- vs. low-heat US in alcohol dependent men, while BOLD reactivity in posterior insula towards the high- vs. low-intensity US itself was increased in patients compared to controls. Recently, Munich et al. (13) showed attenuated amygdala involvement during threat conditioning using mild electric stimulation as US in patients with AUD compared to healthy participants. In spite of general blunting, remaining amygdala activation scaled positively with dependence severity, as well as measures of depression, anxiety, and perceived stress (13). While subjective (12) or physiological conditioned responses (13) did not differ between AUD patients and healthy controls in these imaging studies, two laboratory studies showed blunted

differential physiological responses during Pavlovian threat conditioning in high- compared to low-risk AUD populations (14,15). In line with these findings, reduced amygdala activation has further been observed in response to aversion-inducing, alcohol-related cues in patients with AUD compared to control participants (16).

On the other hand, generic or idiosyncratically appetitive conditioned cues like the sight or smell of an alcoholic beverage have been shown to bias attention and approach tendencies, induce physiological arousal, and often increase subjective craving in AUD (e.g., 17,18), for review, see 19). BOLD responses elicited by such alcohol-associated cues were predictive of subsequent relapse, most consistently in the VS (20–22). At the same time, systematic investigations of the underlying acquisition process of drug conditioning in AUD are sparse. Mayo et al. (23) showed that a novel cue paired with alcohol elicited increased orienting responses that correlated with subjective liking of alcohol in social drinkers. In another study, only participants scoring low on self-reported alcohol sensitivity – a proposed risk phenotype for AUD (24) – demonstrated conditioned neurophysiological responses during second-order conditioning with an alcoholic olfactory cue, suggesting this group might be more susceptible to attribute incentive salience to novel, alcohol-associated cues (25). In addition, we previously showed an increased ability of de-novo conditioned Pavlovian cues to bias instrumental choice behavior in recently detoxified alcohol dependent patients compared to healthy participants, measured using a Pavlovian-to-Instrumental transfer (PIT) task (26–28). Moreover, PIT-related neural activity in the NAcc was increased in prospective relapsers (26,28).

Altogether, impaired threat conditioning in combination with increased cue reactivity could point towards a unique pattern of associative learning alterations in AUD. On the one hand, reward-associated Pavlovian conditioning might be exaggerated, resulting in elevated reactivity towards drug-associated cues. On the other, a reduction in threat conditioning could make subjects more vulnerable to engage in drug-taking behaviors despite severe negative consequences (29). To test this hypothesis, we here investigate for the first time appetitive and aversive de-novo conditioning as part of a PIT paradigm during fMRI in a large sample of

62 recently detoxified AUD patients and 63 matched control participants. We further explore the behavioral and clinical relevance of these associative learning processes by linking differential BOLD responses during Pavlovian learning to the instrumental choice bias in the subsequent PIT phase and prospective relapse risk during a 12-month follow-up period.

2. Methods and Materials

2.1 Participants

As a part of the LeAD study (Learning and Relapse Risk in Alcohol Dependence; clinical trial preregistration identifier: NCT01679145), 62 recently detoxified alcohol-dependent patients (referred to hereinafter as AUD) and 63 healthy controls matched for age, gender, and smoking status were included at two German study sites in Berlin and Dresden (Table 1; see Supplementary Material for exclusion criteria including Table S1 and Figure S1 for participant flowchart). Only participants showing a significant degree of CS-US contingency knowledge post-learning were included in the final analyses (Figure 1B). After detoxification, patients were followed up for twelve months to assess relapse status (see Supplementary Material for details on follow-up assessments). Follow-up information were available for 44 AUD patients (27 relapsers vs. 17 abstainers).

2.2.1 PIT Paradigm

The paradigm consists of four parts: instrumental conditioning, Pavlovian conditioning, Pavlovian-to-Instrumental Transfer (PIT), and a forced-choice task to assess CS-US contingency awareness (see Supplementary Material and Figure S2 for task details).

Instrumental conditioning. Participants learned to collect ’good‘ shells and to leave ’bad‘ shells via probabilistic monetary feedback. Shells could be collected via repeated button presses and participants completed up to 120 trials depending on task performance.

Pavlovian conditioning. The task consisted of two appetitive conditions (CS paired with monetary win +2€ or +1€, respectively), two aversive conditions (CS followed by monetary loss -1€ or -2€, respectively), and a neutral control condition without monetary feedback (0€), using five different multimodal cues as CSs (see Figure 1A). Each CS was presented 16 times, resulting in a total of 80 trials. Participants were instructed to attend to the relations between CS and US and to memorize the pairs. They were further informed that they would receive the displayed, cumulated money after the session.

Pavlovian-to-Instrumental Transfer. During the PIT phase the influence of the learned Pavlovian conditioned stimuli on instrumental choice behavior was measured. Participants performed the instrumental task while one of the Pavlovian CSs tiled the background without receiving feedback.

Forced-choice task. Finally, CS-US contingency knowledge of Pavlovian learning was assessed, where participants had to choose the higher-valued CS out of two CSs presented on the left and right site of the screen. Each CS combination was presented 3 times in pseudo-randomized order. Only participants performing significantly over chance (83% of AUD and 91% of control participants) were considered contingency aware and included in the final analyses, as contingency awareness seems necessary for Pavlovian trace conditioning to occur (30–32). Likewise, PIT effects can only be meaningfully analyzed in contingency aware participants (26–28) (Figure 1B; see Supplementary Material and Table S2 for sample characteristics of aware vs. unaware participants).

2.4 Data analysis

Behavioral data were analyzed using Matlab R2019b (The MathWorks, Inc., Natick, Massachusetts, United States) and R version 3.6.1 (33). The alpha level was set at $p<.05$ for all analyses.

CS-US contingency awareness. Contingency awareness was measured as percentage of higher-valued CS choices during the forced-choice task and group differences were examined via Mann-Whitney U test (see Supplementary Material for more detailed analyses).

Pleasantness and arousal ratings. Subjective ratings of CS pleasantness and arousal, obtained at the end of the PIT paradigm, were analyzed in separate linear mixed-effects models (LMMs) including CS value, group, and study site (see Supplementary Material for details). Aversive and appetitive conditioning were investigated separately, given first evidence of deficits in Pavlovian threat conditioning in high-risk samples (14,15) and attenuated neural differentiation in AUD (12,13), while lacking systematic investigations on

appetitive conditioning in AUD.

Behavioral PIT effect. The behavioral PIT effect was analyzed as previously described (26) (see Supplementary Material).

fMRI. After standardized preprocessing (see Supplementary Material), an event-related analysis was applied using the GLM approach within SPM 12 (Welcome Department of Imaging Neuroscience; www.fil.ion.ucl.ac.uk/spm/) on two levels. For each participant, onset regressors for each CS and US type were modeled as stick functions and convolved with the canonical HRF. Additional nuisance regressors included an eye-tracker recalibration period after half of the trials (mean duration 71.6 s), modeled as box-car function and the 6 movement parameters to account for movement-related variance. Baseline contrasts for each CS were computed and entered into a random-effects flexible factorial model on the second level, together with the group factor (AUD/HC). We investigated main effects across participants as well as group differences for the following three contrasts: Pavlovian learning was probed by contrasting CSs across valence conditions with the control condition (0€), taking into account the grading within appetitive and aversive conditions (i.e. Pavlovian CSs: -2€ -1€ 0€ +1€ +2€, contrast ‚Pavlovian learning': [+2 +1 -6 +1 +2]). We then separately investigated appetitive and aversive Pavlovian learning (contrast ‚aversive Pavlovian learning': [+2 +1 -3 0 0]; contrast ‚appetitive Pavlovian learning': [0 0 -3 +1 +2]). Group differences were investigated by testing the group x contrast interaction, followed by post-hoc t-tests in case of a significant effect.

We focused our analyses on three predefined regions-of-interest (ROI): Amygdala and hippocampus, due to their central role in appetitive and aversive Pavlovian (trace) conditioning (8,34–36) as well as the ventral striatum (VS) (10,37), critically involved in reward processing (38) and previously shown to modulate PIT effects in AUD (26,28,39). Bilateral ROIs for amygdala and hippocampus were derived using the WFU PickAtlas (http://www.fmri.wfubmc.edu/download.html) and the VS as a functionally defined mask using the BrainMap database (40) similar to previous publications (41,42). ROI-analyses were performed at $p<0.05$ FWE-correction, complemented by exploratory whole-brain analyses at

p<0.05 FWE correction at the cluster level, using a cluster-forming threshold of p<0.001 uncorrected and cluster extend of ten contiguous voxels. To account for multiple comparisons across ROIs, p-values were additionally adjusted for the number of ROIs using Bonferroni correction.

Brain-behavior associations: Individual PIT effects (see Supplementary Materials) were entered as a covariate within SPM in a separate second-level GLM with the ’Pavlovian learning‘ contrast and the group factor (AUD/HC), allowing for an interaction between group and covariate. We focused our ROI analysis on the amygdala and VS shown to modulate neural PIT effects (43–45).

To investigate whether neural signatures during Pavlovian learning were predictive of subsequent relapse, we re-run the flexible-factorial model and informed the group factor by patients’ prospective relapse status (relapsers vs. abstainers vs. HC). We further explored whether neural responses during Pavlovian learning correlated with relapse latency in prospective relapsers using simple regression analysis with the ‘Pavlovian learning‘ contrast and the number of abstinence days till relapse as a covariate.

Study site was included as additional covariate in all analyses.

Results

Explicit learning of CS-US associations: contingency awareness

Contingency awareness was assessed post-learning in a forced-choice task, using data from all participants providing high-quality fMRI data (75 AUD patients vs. 69 HC, Figure 1B, see Supplementary Figure S1 for participant flowchart). Overall performance was at 86.6% correct choices (SD=17.4; range: 16.7-100), with no differences between groups (W=2515.5, p=.77), indicating equal levels of contingency awareness (Figure 1B; see also Supplementary Figure S4). All subsequent analyses are based on participants performing significantly over chance (i.e., 'Pavlovian learner', as confirmed by binomial test).

Subjective measures of Pavlovian learning: pleasantness and arousal ratings

Subjective CS pleasantness and arousal ratings, acquired post conditioning, were significantly influenced by the conditioning protocol, evident in a linear effect of CS value on pleasantness ratings (b=0.15, SE=0.06, t=2.78, p=.006) and a linear and quadratic effect on subjective arousal (b_{linear}=0.09, SE=0.04, t=2.22, p=.027; $b_{quadratic}$=0.08, SE=0.04, t=2.24, p=.026; Supplementary Figure S5). This indicated that participants' pleasantness and arousal ratings reflected Pavlovian value after conditioning. Arousal ratings were higher in AUD patients compared to controls across cues (b=-0.58, SE=0.26, t=-2.22, p=.028), but we did not observe a group by value interaction, indicating groups did not differ in conditioned responses (pleasantness: p=.358; arousal: p≥.158). Separate investigation of appetitive and aversive conditioning revealed the observed behavioral effects were driven by appetitive CSs (pleasantness: b=0.26, SE=0.12, t=2.21, p=.028; arousal: b=0.22, SE=.09, t=2.46, p=.015) rather than aversive CSs (pleasantness and arousal p>=.386), without significant effects of group or group by CS value interaction in neither analysis (p≥.118).

Neural representation of Pavlovian learning: BOLD signals towards appetitive and aversive Pavlovian cues

Across participants, Pavlovian learning induced marginally increased BOLD

responses in right amygdala ($p_{FWE\ ROI}$ =.099; Table 2). Separate investigation of appetitive and aversive Pavlovian conditioning revealed significantly increased BOLD responses towards reward-predicting cues in the left VS ($_{pFWE\ ROI}$=.05; Table 2), while aversive Pavlovian conditioning showed no significant differential BOLD responses. No additional activated clusters survived in the whole-brain analyses.

Group comparison revealed significant different engagement of right amygdala during Pavlovian conditioning (amygdala right: [x:28, y:-4, z:-22], $F_{1,492}$=14.65, $p_{FWE\ ROI}$=.029). Post-hoc analysis showed that AUD patients exhibited significantly stronger differential BOLD responses in bilateral amygdala towards Pavlovian cues relative to healthy controls (see Table 2; Figure 2; complementary analyses are provided in the Supplementary Material). Investigating differential BOLD responses for appetitive and aversive Pavlovian conditioning separately revealed that the observed group difference was specific for reward-predicting cues, assessed with the appetitive Pavlovian conditioning contrast (amygdala right: [x:26, y:-6, z:-22], $F_{1,492}$=16.75, $p_{FWE\ ROI}$=.006; amygdala left: [x:-24, y:-8, z:-22], $F_{1,492}$=12.84, $p_{FWE\ ROI}$=.045). Here, AUD patients additionally showed stronger recruitment of an anterior cluster within the hippocampus ([x:26, y:-10, z:-22], $F_{1,492}$=14.38, $p_{FWE\ ROI}$=.027; Table 2). In contrast, no group differences emerged during aversive Pavlovian conditioning. Results remained significant when contingency unaware participants were also included (see Supplementary Material).

Association of Pavlovian conditioning with instrumental PIT behavior and prospective relapse

We further investigated whether neural responses during Pavlovian learning were related to the ability of Pavlovian cues to bias subsequent choice behavior (i.e., PIT effect; see Supplementary Table S3 and Figure S3), and to prospective relapse risk.

Across groups, this analysis revealed that increased conditioning-related BOLD activity in right VS was associated with a stronger instrumental choice bias during the subsequent PIT phase ([x:4, y:-14, z:-8], Z=3.48, $p_{FWE\ ROI}$=.05; Supplementary Figure S6).

Group comparisons showed that BOLD activity in left amygdala was predominantly predictive of patients' subsequent choice bias, in contrast to healthy controls (left: [x:-26, y:-2, z:-24], Z=3.35, $p_{FWE\ ROI}$=.048; Figure 3).

Finally, we assessed whether neural signals during *de-novo* conditioning were associated with prospective relapse at 1-year follow-up. Contrasting prospective relapsers with abstainers as well as healthy controls revealed a main effect of group in the right amygdala during Pavlovian learning ([x:26, y:-6, z:-20], $F_{2,416}$=8.58, $p_{FWE\ ROI}$=.033; Figure 4). Post-hoc analyses confirmed that both patient groups showed increased amygdala activity relative to healthy controls (relapser > HC: [x:26, y:-6, z:-20], Z=3.47, $p_{FWE\ ROI}$=.033; abstainer > HC: [x:24, y:-4, z:-20], Z=3.40, $p_{FWE\ ROI}$=.042). Although amygdala activation did not differ between prospective relapsers and abstainers ($p_{FWE\ ROI}$=.22), within patients who relapsed, increased right amygdala activation during Pavlovian learning was associated with reduced relapse latency ([x:20, y:0, z:-20], Z=2.94, $p_{FWE\ ROI}$=.047; Figure 4).

Discussion

In this study, we investigated de-novo Pavlovian conditioning of both appetitive and aversive associations in recently detoxified AUD patients and healthy participants during functional magnetic resonance imaging. While both patients and healthy participants were equally likely to acquire the different CS–US associations in terms of explicit contingency knowledge, Pavlovian CSs elicited significantly stronger BOLD responses in bilateral amygdala in patients compared to controls. This difference was most pronounced for reward-predicting cues. We further related BOLD responses during Pavlovian conditioning to the behavioral choice bias induced by these cues in a subsequent PIT test, as well as to relapse during a 12-month follow-up period. In contrast to healthy participants, left amygdala activation during Pavlovian conditioning was positively associated with the subsequent behavioral PIT effect in patients with AUD, and among patients who relapsed, right amygdala activation was predictive of relapse latency in an exploratory analysis.

Patients with AUD showed elevated amygdala activation towards Pavlovian cues during de-novo conditioning

We observed significant group differences in the amygdala during de-novo Pavlovian conditioning, with stronger differential BOLD responses in patients compared to healthy participants.

Converging lines of evidence identified the amygdala as a core region subserving appetitive and aversive Pavlovian conditioning: The amygdala is critically involved in encoding the state value of motivational salient stimuli, forming CS-US associations, and expression of conditioned responses (34,46–48). Amygdala responsivity has been shown to capture individual differences in human threat conditioning, as BOLD signals in this region correlate with physiological conditioning indices (49–52). Furthermore, amygdala activation plays a vital role during Pavlovian relapse effects, i.e. the return of conditioned responses

after extinction (53,54), highlighting the importance of this structure for acquisition, recall and expression of conditioned responses.

Our observation of elevated differential amygdala activation during Pavlovian conditioning in AUD patients compared to healthy participants therefore likely reflects enhanced neural encoding of Pavlovian associations – especially rewarding ones – and could reflect greater susceptibility to assign incentive salience to novel, reward-related cues in AUD (1). To our knowledge, this is the first study investigating appetitive Pavlovian de-novo conditioning in AUD patients. Further evidence for enhanced drug-related Pavlovian learning in at-risk participants comes from Fleming and colleagues (25), where Pavlovian de-novo conditioning using an alcoholic olfactory cue only induced subjective craving and conditioned event-related potentials in low but not high alcohol sensitive participants, a phenotype associated with risk to develop AUD (24).

Regarding aversive Pavlovian conditioning, Muench et al. (13) observed overall attenuated differential amygdala activation during de-novo Pavlovian threat conditioning in AUD patients compared to healthy participants. Interestingly, however, differential amygdala responses scaled positively with AUD symptom severity (13). During instructed threat conditioning, where CS-US contingencies are known in advance, alcohol-dependent men showed attenuated differential BOLD responses towards a high- vs. low-heat predicting cue in cortical regions associated with negative affect regulation, including the pregenual ACC and medial PFC, together with increased posterior insula activation towards the high- vs. low-intensity US itself (12). Further evidence for altered threat conditioning in AUD comes from two studies in high-risk populations (14,15). Finn et al. (15) found that men with a high family history of AUD compared to men without such family history failed to acquire differential SCRs towards threat compared to neutral cues due to reduced CS+ responsiveness. Attenuated differential SCRs and startle responses towards aversive vs. neutral CSs were also observed in young binge drinkers compared to non-binge drinkers (13), also corroborating rat studies showing that multiple episodes of ethanol withdrawal can impair fear conditioning due to lower CS+ responsiveness (55,56).

In our study, investigating aversive and appetitive Pavlovian conditioning separately showed that the group difference in amygdala was primarily driven by enhanced BOLD responses towards reward- and not loss-predicting cues. However, we would refrain from drawing specific conclusions about aversive conditioning in AUD, as the aversive contrast in our paradigm did not elicit significant differential BOLD activation across participants (see also (35)). Therefore, cues signaling threat like electric shock or loud noise might be better suited to study aversive associative learning in future studies.

Conditioned amygdala responses are related to Pavlovian-to-Instrumental transfer and relapse latency in patients with AUD

The PIT paradigm enables investigation of the influence of Pavlovian cues on instrumental behavior - an effect called Pavlovian-to-Instrumental transfer (27,28,57). PIT effects are mediated by distinct regions within the NAcc and amygdala (44,58,59) and have been discussed as a potential mechanism contributing to habit formation and habitual drug use in AUD (3,60,61).

By relating neural activity during Pavlovian conditioning to the subsequent behavioral PIT effect, we showed that ventral striatal BOLD responses were positively correlated with the strength of the PIT effect in both patients and control participants. Amygdala activation during Pavlovian conditioning significantly correlated with instrumental choices during PIT in patients with AUD compared to healthy controls. This observation suggests that amygdala engagement during Pavlovian conditioning contributes to instrumental choices towards these Pavlovian cues, and that this association is pronounced in patients with AUD, underlining the behavioral relevance of our neural finding.

Previous research showed that BOLD responses in the NAcc during the PIT phase were predictive of subsequent relapse during follow-up in AUD patients (26,28). Therefore, we assessed whether the neural signatures of Pavlovian conditioning represent a potential marker for prospective relapse within a 12-month follow-up period. Both prospective

relapsers and abstainers showed elevated amygdala responses during Pavlovian conditioning compared to healthy participants, and patient groups did not differ significantly in overall amygdala reactivity. However, we found a significant inverse correlation between right amygdala activation in response to Pavlovian CSs and relapse latency in prospectively relapsing patients. This study, if replicated, may suggest that increased amygdala reactivity towards Pavlovian cues is not a general risk factor of AUD, but could decrease relapse latency in vulnerable persons. Patients who abstained might have additional protective factors, helping them to stay abstinent despite increased amygdala reactivity during Pavlovian learning. Cue-reactivity research revealed that abstinent compared to non-abstinent AUD patients showed increased functional connectivity between limbic regions and prefrontal areas in a cue-reactivity paradigm, potentially helping them to stay abstinent in the presence of craving-inducing alcohol-cues (62). However, increased cue-induced limbic brain activation may not simply promote relapse, but could contribute to salience attribution also required for inhibitory control (20,63). Complex top-down and bottom-up mechanisms might constitute an important moderating factor also shown to critically interact with cue-reactivity in AUD (64,65).

Limitations

Several limitations need to be considered: First, prospective studies in participants at risk are needed to elucidate whether the observed alterations in Pavlovian conditioning represent a predisposing factor for AUD, or rather develop throughout the disease. Furthermore, our paradigm might not be optimal to disentangle appetitive and aversive conditioning, as the aversive contrast failed to significantly engage relevant brain structures. Third, we acquired no additional psychophysiological measure of conditioned responding, e.g. skin conductance, limiting comparability between studies. Fourth, the observed group difference in amygdala activation during Pavlovian conditioning was due to both increased BOLD responses towards Pavlovian cues in AUD, as well as towards the neutral cue in healthy participants. Our control

condition might be affectively more ambiguous compared to a neutral cue not paired with any outcome, as often used in (fear) conditioning paradigms, highlighting the need for careful consideration of adequate baseline conditions in future studies. Moreover, we did not investigate conditioning of drug-related cues, which might tap into more disease specific mechanisms within largely overlapping neural circuits (66,67).

Conclusion and future directions

To conclude, we provide evidence for altered Pavlovian learning processes in patients with AUD, reflected in increased amygdala recruitment that was especially pronounced during reward-associative learning. Increased amygdala reactivity was related to subsequent PIT behavior as well as to relapse latency during a 12-month follow-up period. These findings may reflect greater susceptibility to assign incentive salience to novel, reward-related cues in AUD (1), a process that might contribute to bias patients' behavioral choices in the presence of these Pavlovian stimuli.

Our findings extend evidence in AUD and related high-risk populations on Pavlovian conditioning, and point towards patterns of associative learning alterations, whereby conditioned responses towards reward- or drug-associated Pavlovian cues are increased (25), while learning from threat-signaling cues is abolished (12–15). This might promote elevated reactivity towards reward-associated cues including drug cues on the one hand, while subjects engage in conditioned behaviors despite severe negative consequences on the other.

Interestingly, the reported conditioning alterations in AUD are different from that seen in other patient populations including post-traumatic stress disorder and anxiety, whereby both threat and safety cues elicit increased physiological responses and neural activation of the amygdala, suggesting abnormal fear generalization (68,69). Given the outlined evidence, investigating both reward and threat conditioning processes in mental disorders could represent a fruitful avenue for future research, as it enables to dissociate learning alterations

in different value domains (70,71). Moreover, investigating individual differences in these learning mechanisms might provide valuable insights about the role of Pavlovian conditioning in addiction maintenance (e.g., (72)).

Ultimately, characterizing alterations in neural structures subserving Pavlovian learning processes, that is, mechanisms at the center of influential theories of addiction (1–4), could foster our understanding of AUD as a disorder driven by maladaptive learning and provide targets for future therapeutic interventions aim to counteract the motivational power of alcohol-related cues (73).

Acknowledgments

We thank the LeAD study teams in Dresden and Berlin for data acquisition. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) FOR 1617 – grants HE2597/14-1, HE2597/14-2, SCHL1969/2-2, ZI1119/3-1, ZI1119/3-2, ZI1119/4-1 – and in part by TRR 265, Project ID 402170461, subprojects B01 and B03 (57) and the BMBF (grant 01ZX1311H). FS was supported by the Heisenberg Program (SCHL 1969/5-1). QJMH acknowledges support by the UCLH NIH BRC.

Disclosures

QJMH has obtained fees and options for consultancies for Aya Technologies and Alto Neuroscience. AH, CE, FS, MG, MNS, MS, KC, and USZ reported no biomedical financial interests or potential conflicts of interest.

References

1. Berridge KC, Robinson TE (2016): Liking, wanting, and the incentive-sensitization theory of addiction. *Am Psychol* 71: 670–679.
2. Milton AL, Everitt BJ (2012): The persistence of maladaptive memory: Addiction, drug memories and anti-relapse treatments. *Neurosci Biobehav Rev* 36: 1119–1139.
3. Heinz A, Beck A, Halil MG, Pilhatsch M, Smolka MN, Liu S (2019): Addiction as learned behavior patterns. *J Clin Med* 8: 1–9.
4. Robbins TW, Ersche KD, Everitt BJ (2008): Drug addiction and the memory systems of the brain. *Ann N Y Acad Sci* 1141: 1–21.
5. Robinson TE, Berridge KC (1993): The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res Rev* 18: 247–291.
6. Doñamayor N, Ebrahimi C, Garbusow M, Wedemeyer F, Schlagenhauf F, Heinz A (2021): Instrumental and Pavlovian Mechanisms in Alcohol Use Disorder. *Curr Addict Reports* 8: 156–180.
7. Martin-Soelch C, Linthicum J, Ernst M (2007): Appetitive conditioning: Neural bases and implications for psychopathology. *Neurosci Biobehav Rev* 31: 426–440.
8. Sehlmeyer C, Schöning S, Zwitserlood P, Pfleiderer B, Kircher T, Arolt V, Konrad C (2009): Human fear conditioning and extinction in neuroimaging: a systematic review. *PLoS One* 4: e5865.
9. Tapia León I, Kruse O, Stalder T, Stark R, Klucken T (2018): Neural correlates of subjective CS/UCS association in appetitive conditioning. *Hum Brain Mapp* 39: 1637–1646.
10. Fullana MA, Harrison BJ, Soriano-Mas C, Vervliet B, Cardoner N, Àvila-Parcet A, Radua J (2016): Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry* 21: 500–508.
11. Pietrock C, Ebrahimi C, Katthagen TM, Koch SP, Heinz A, Rothkirch M, Schlagenhauf F (2019): Pupil dilation as an implicit measure of appetitive Pavlovian learning. *Psychophysiology* 56: 00:e13463.
12. Yang H, Devous MD, Briggs RW, Spence JS, Xiao H, Kreyling N, Adinoff B (2013): Altered Neural Processing of Threat in Alcohol-Dependent Men. *Alcohol Clin Exp Res* 37: 2029–2038.
13. Muench C, Charlet K, Balderston NL, Grillon C, Heilig M, Cortes CR, *et al.* (2021): Fear conditioning and extinction in alcohol dependence: Evidence for abnormal amygdala reactivity. *Addict Biol* 26: 1–12.
14. Stephens DN, Ripley TL, Borlikova G, Schubert M, Albrecht D, Hogarth L, Duka T (2005): Repeated ethanol exposure and withdrawal impairs human fear conditioning and

depresses long-term potentiation in rat amygdala and hippocampus. *Biol Psychiatry* 58: 392–400.

15. Finn PR, Kessler DN, Hussong AM (1994): Risk for alcoholism and classical conditioning to signals for punishment: Evidence for a weak behavioral inhibition system. *J Abnorm Psychol* 103: 293–301.
16. Kim SM, Han DH, Min KJ, Kim BN, Cheong JH (2014): Brain activation in response to craving- and aversion-inducing cues related to alcohol in patients with alcohol dependence. *Drug Alcohol Depend* 141: 124–131.
17. Wrase J, Schlagenhauf F, Kienast T, Wüstenberg T, Bermpohl F, Kahnt T, *et al.* (2007): Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *Neuroimage* 35: 787–794.
18. Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J (2011): Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychol Sci* 22: 490–497.
19. Cofresí RU, Bartholow BD, Piasecki TM (2019): Evidence for incentive salience sensitization as a pathway to alcohol use disorder. *Neurosci Biobehav Rev* 107: 897–926.
20. Beck A, Wüstenberg T, Genauck A, Wrase J, Schlagenhauf F, Smolka MN, *et al.* (2012): Effect of Brain Structure, Brain Function, and Brain Connectivity on Relapse in Alcohol-Dependent Patients. *Arch Gen Psychiatry* 69: 842.
21. Grüsser SM, Wrase J, Klein S, Hermann D, Smolka MN, Ruf M, *et al.* (2004): Cue-induced activation of the striatum and medial prefrontal cortex is associated with subsequent relapse in abstinent alcoholics. *Psychopharmacology (Berl)* 175: 296–302.
22. Reinhard I, Leménager T, Fauth-Bühler M, Hermann D, Hoffmann S, Heinz A, *et al.* (2015): A comparison of region-of-interest measures for extracting whole brain data using survival analysis in alcoholism as an example. *J Neurosci Methods* 242: 58–64.
23. Mayo LM, de Wit H (2016): Acquisition of conditioned responses to a novel alcohol-paired cue in social drinkers. *J Stud Alcohol Drugs* 77: 317–326.
24. Schuckit MA (1994): Low level of response to alcohol as a predictor of future alcoholism. *Am J Psychiatry* 151: 184–189.
25. Fleming KA, Cofresí RU, Bartholow BD (2021): Transfer of incentive salience from a first-order alcohol cue to a novel second-order alcohol cue among individuals at risk for alcohol use disorder: electrophysiological evidence. *Addiction* 116: 1734–1746.
26. Chen K, Schlagenhauf F, Sebold M, Kuitunen-Paul S, Chen H, Huys QJM, *et al.* (2022): The Association of Non–Drug-Related Pavlovian-to-Instrumental Transfer Effect in Nucleus Accumbens With Relapse in Alcohol Dependence: A Replication. *Biol Psychiatry*. https://doi.org/10.1016/j.biopsych.2022.09.017

27. Garbusow M, Schad DJ, Sommer C, Jünger E, Sebold M, Friedel E, *et al.* (2014): Pavlovian-to-instrumental transfer in alcohol dependence: A pilot study. *Neuropsychobiology* 70: 111–121.
28. Garbusow M, Schad DJ, Sebold M, Friedel E, Bernhardt N, Koch SP, *et al.* (2016): Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addict Biol* 21: 719–731.
29. Nees F, Heinrich A, Flor H (2015): A mechanism-oriented approach to psychopathology: The role of Pavlovian conditioning. *Int J Psychophysiol* 98: 351–364.
30. Clark RE, Squire LR (1998): Classical Conditioning and Brain Systems: The Role of Awareness. *Science (80-)* 280: 77–81.
31. Knight DC, Cheng DT, Smith CN, Stein E a, Helmstetter FJ (2004): Neural substrates mediating human delay and trace fear conditioning. *J Neurosci* 24: 218–228.
32. Weike AI, Schupp HT, Hamm AO (2007): Fear acquisition requires awareness in trace but not delay conditioning. *Psychophysiology* 44: 170–80.
33. R Core Team (2017): R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org
34. Büchel C, Dolan RJ (2000): Classical fear conditioning in functional neuroimaging. *Curr Opin Neurobiol* 10: 219–23.
35. Ebrahimi C, Koch SP, Friedel E, Crespo I, Fydrich T, Ströhle A, *et al.* (2017): Combining D-cycloserine with appetitive extinction learning modulates amygdala activity during recall. *Neurobiol Learn Mem* 142: 209–217.
36. Büchel C, Dolan RJ, Armony JL, Friston KJ (1999): Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *J Neurosci* 19: 10869–10876.
37. Gottfried JA, O'Doherty J, Dolan RJ (2002): Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J Neurosci* 22: 10829–37.
38. Sescousse G, Caldú X, Segura B, Dreher J-C (2013): Processing of primary and secondary rewards: A quantitative meta-analysis and review of human functional neuroimaging studies. *Neurosci Biobehav Rev* 37: 681–696.
39. Schad DJ, Garbusow M, Friedel E, Sommer C, Sebold M, Hägele C, *et al.* (2018): Neural correlates of instrumental responding in the context of alcohol-related cues index disorder severity and relapse risk. *Eur Arch Psychiatry Clin Neurosci* 269: 295–308.
40. Nielsen FA, Hansen LK (2002): Automatic anatomical labeling of Talairach coordinates and generation of volumes of interest via the BrainMap database. *Neuroimage* 16: 2–6.
41. Chen H, Mojtahedzadeh N, Belanger MJ, Nebe S, Kuitunen-Paul S, Sebold M, *et al.* (2021): Model-Based and Model-Free Control Predicts Alcohol Consumption

Developmental Trajectory in Young Adults: A 3-Year Prospective Study. *Biol Psychiatry* 89: 980–989.

42. Nebe S, Kroemer NB, Schad DJ, Bernhardt N, Sebold M, Müller DK, *et al.* (2018): No association of goal-directed and habitual control with alcohol consumption in young adults. *Addict Biol* 23: 379–393.
43. Geurts DEM, Huys QJM, den Ouden HEM, Cools R (2013): Aversive Pavlovian Control of Instrumental Behavior in Humans. *J Cogn Neurosci* 25: 1428–1441.
44. Prevost C, Liljeholm M, Tyszka JM, O'Doherty JP (2012): Neural Correlates of Specific and General Pavlovian-to-Instrumental Transfer within Human Amygdalar Subregions: A High-Resolution fMRI Study. *J Neurosci* 32: 8383–8390.
45. Talmi D, Seymour B, Dayan P, Dolan RJ (2008): Human pavlovian-instrumental transfer. *J Neurosci* 28: 360–368.
46. LeDoux JE (2000): Emotion Circuits in the Brain. *Annu Rev Neurosci* 23: 155–184.
47. Greco JA, Liberzon I (2016): Neuroimaging of fear-associated learning. *Neuropsychopharmacology* 41: 320–334.
48. Janak PH, Tye KM (2015): From circuits to behaviour in the amygdala. *Nature* 517: 284–292.
49. LaBar KS, Gatenby JC, Gore JC, LeDoux JE, Phelps EA (1998): Human Amygdala Activation during Conditioned Fear Acquisition and Extinction: a Mixed-Trial fMRI Study. *Neuron* 20: 937–945.
50. Cheng DT, Knight DC, Smith CN, Helmstetter FJ (2006): Human amygdala activity during the expression of fear responses. *Behav Neurosci* 120: 1187–1195.
51. MacNamara A, Rabinak CA, Fitzgerald DA, Zhou XJ, Shankman SA, Milad MR, Phan KL (2015): Neural correlates of individual differences in fear learning. *Behav Brain Res* 287: 34–41.
52. Van Well S, Visser RM, Scholte HS, Kindt M (2012): Neural substrates of individual differences in human fear learning: Evidence from concurrent fMRI, fear-potentiated startle, and US-expectancy data. *Cogn Affect Behav Neurosci* 12: 499–512.
53. Ebrahimi C, Koch SP, Pietrock C, Fydrich T, Heinz A, Schlagenhauf F (2019): Opposing roles for amygdala and vmPFC in the return of appetitive conditioned responses in humans. *Transl Psychiatry* 9: 148.
54. Lonsdorf TB, Haaker J, Kalisch R (2014): Long-term expression of human contextual fear and extinction memories involves amygdala, hippocampus and ventromedial prefrontal cortex: A reinstatement study in two independent samples. *Soc Cogn Affect Neurosci* 9: 1973–1983.
55. Stephens DN, Brown G, Duka T, Ripley TL (2001): Impaired fear conditioning but enhanced seizure sensitivity in rats given repeated experience of withdrawal from

alcohol. *Eur J Neurosci* 14: 2023–2031.
56. Ripley TL, O'Shea M, Stephens DN (2003): Repeated withdrawal from ethanol impairs acquisition but not expression of conditioned fear. *Eur J Neurosci* 18: 441–448.
57. Cartoni E, Balleine B, Baldassarre G (2016): Appetitive Pavlovian-instrumental Transfer: A review. *Neurosci Biobehav Rev* 71: 829–848.
58. Corbit LH, Balleine BW (2005): Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *J Neurosci* 25: 962–970.
59. Corbit LH, Balleine BW (2011): The general and outcome-specific forms of pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *J Neurosci* 31: 11786–11794.
60. Garbusow M, Nebe S, Sommer C, Kuitunen-Paul S, Sebold M, Schad DJ, *et al.* (2019): Pavlovian-To-Instrumental Transfer and Alcohol Consumption in Young Male Social Drinkers: Behavioral, Neural and Polygenic Correlates. *J Clin Med* 8: 1188.
61. Corbit LH, Janak PH (2016): Habitual Alcohol Seeking: Neural Bases and Possible Relations to Alcohol Use Disorders. *Alcohol Clin Exp Res* 40: 1380–1389.
62. Strosche A, Zhang X, Kirsch M, Hermann D, Ende G, Kiefer F, Vollstädt-Klein S (2021): Investigation of brain functional connectivity to assess cognitive control over cue-processing in Alcohol Use Disorder. *Addict Biol* 26: 1–13.
63. Sommer C, Garbusow M, Jünger E, Pooseh S, Bernhardt N, Birkenstock J, *et al.* (2017): Strong seduction: Impulsivity and the impact of contextual cues on instrumental behavior in alcohol dependence. *Transl Psychiatry* 7. https://doi.org/10.1038/tp.2017.158
64. Stuke H, Gutwinski S, Wiers CE, Schmidt TT, Gröpper S, Parnack J, *et al.* (2016): To drink or not to drink: Harmful drinking is associated with hyperactivation of reward areas rather than hypoactivation of control areas in men. *J Psychiatry Neurosci* 41: E24–E36.
65. Vollstädt-Klein S, Gerhardt S, Lee A, Strosche A, Sharafi G, Nuriyeva R, *et al.* (2020): Interaction between behavioral inhibition and neural alcohol cue-reactivity in ADHD and alcohol use disorder. *Psychopharmacology (Berl)* 237: 1691–1707.
66. Hill-Bowen LD, Riedel MC, Poudel R, Salo T, Flannery JS, Camilleri JA, *et al.* (2021): The cue-reactivity paradigm: An ensemble of networks driving attention and cognition when viewing drug and natural reward-related stimuli. *Neurosci Biobehav Rev* 130: 201–213.
67. Tang DW, Fellows LK, Small DM, Dagher A (2012): Food and drug cues activate similar brain regions: A meta-analysis of functional MRI studies. *Physiol Behav* 106: 317–324.
68. Duits P, Cath DC, Lissek S, Hox JJ, Hamm AO, Engelhard IM, *et al.* (2015): Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depress Anxiety* 32: 239–253.
69. Pittig A, Treanor M, LeBeau RT, Craske MG (2018): The role of associative fear and

644 avoidance learning in anxiety disorders: Gaps and directions for future research.
645 *Neurosci Biobehav Rev* 88: 117–140.
646 70. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, *et al.* (2010): Research
647 Domain Criteria (RDoC): Toward a new classification framework for research on mental
648 disorders. *American Journal of Psychiatry*, vol. 167. pp 748–751.
649 71. Heinz A, Schlagenhauf F, Beck A, Wackerhagen C (2016): Dimensional psychiatry:
650 mental disorders as dysfunctions of basic learning mechanisms. *J Neural Transm* 123:
651 809–821.
652 72. Burnette EM, Grodin EN, Schacht JP, Ray LA (2021): Clinical and Neural Correlates of
653 Reward and Relief Drinking. *Alcohol Clin Exp Res* 45: 194–203.
654 73. Oberlin BG, Shen YI, Kareken DA (2020): Alcohol Use Disorder Interventions Targeting
655 Brain Sites for Both Conditioned Reward and Delayed Gratification. *Neurotherapeutics*
656 17: 70–86.
657 74. Schmidt LG, Gastpar M, Falkai P, Gäbel W (Eds.) (2006): *Evidenzbasierte Suchtmedizin.*
658 *Behandlungsleitlinie Substanzbezogene Störungen*. Köln: Deutscher Ärzte-Verlag.
659 75. Lehrl S (2005): *Multiple Choice Vocabulary Test- German Version (MWT-B)*. Balingen:
660 Spitta Verlag.
661 76. Reitan RM (1992): *Trial Making Test. Manual for Administration and Scoring*. Tucson, AZ:
662 Reitan Neuropsychology Laboratory.
663 77. Wittchen H-U, Pfister H (1997): *DIA-X Interview. Instruktionsmanual Zur Durchführung*
664 *von DIA-X-Interviews (Instruction Manual for the DIA-X-Interview)*. Frankfurt: Swets &
665 Zeitlinger.
666 78. Mann K, Ackermann K (2000): The OCDS-G: Psychometric characteristics of the German
667 version of the Obsessive Compulsive Drinking Scale. *Sucht* 46: 90–100.
668 79. Meule A, Vögele C, Kübler A (2011): Psychometrische evaluation der Deutschen Barratt
669 Impulsiveness scale - Kurzversion (BIS-15). *Diagnostica* 57: 126–133.
670
671

672 **Table 1**: Sample characteristics

	N	AUD Patients		N	Healthy Controls		p-value
Sociodemographics							
Gender	62	♀:13, ♂:49		63	♀:10, ♂:53		.61[a]
		Mean	SD		Mean	SD	
Age in years	62	43.98	11.59	63	42.86	11.19	.59
Smokers		75.8 %			73.0 %		.88[a]
Education (in years)	62	14.37	3.12	61	15.9	3.89	**.02**
Socio-economic status (SES)	54	-0.41	1.88	42	0.49	1.81	**.02**
Neurocognitive functioning							
Verbal intelligence (MWT-B)	60	104.52	9.43	62	104.66	9.53	.93
TMT-A (seconds)	60	29.42	8.7	62	28.31	9.39	.50
TMT-B (seconds)	60	69.98	26.46	62	60.16	22.54	**.03**
AUD severity							
years with diagnosis (DSM-IV)	57	11.35	10.24		-		
number of DSM-IV symptoms	58	5.71	1.24	63	0.51	1.03	**<.001**
Severity of AUD (ADS)	62	15.31	7.06	63	1.94	2.93	**<.001**
Lifetime alcohol consumption in kg (pure alcohol)[b]	62	1717.26	1180.3	63	303.67	988.74	**<.001**
Craving (OCDS-G total score)	61	12.84	8.33	62	2.87	2.89	**<.01**
days of abstinence before scanning	62	20.31	11.65	63	88.89	342.16	.12
Personality							
Impulsivity (BIS total score)	59	30.47	6.4	62	29.15	5.55	.23

673

674 Socio-economic status (SES) was computed as the sum of self-rated z-transformed scores of social
675 status, household income, and inverse personal dept scores (74). Verbal intelligence was assessed
676 with the Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B; German Multiple-Choice Vocabulary
677 Intelligence Test) (75) and executive functioning by the Trial Making Test A and B (76). Amount of
678 lifetime alcohol intake was measured by the CIDI (77), current craving by the Obsessive Compulsive
679 Drinking Scale (OCDS-G; German version) (78), and trait impulsivity using the Barratt Impulsiveness.

680 Scale 15 (BIS-15; German version) (79). AUD: alcohol use disorder; [a] p-value of χ^2-test, independent

681 t-test otherwise; [b] prior to detoxification in AUD patients.

682

683 **Table 2**: Region-of-interest analyses of Pavlovian conditioning

Analysis	Contrast	Region	Side	k	Peak voxel MNI x	y	z	Z_{max}	p_{FWE}
all participants									
	Pavlovian conditioning	Amygdala	R		28	-2	-14	3.08	0.033
	[$CS_{-2€} > CS_{-1€} > CS_{0€} < CS_{+1€} < CS_{+2€}$]	VS	L		-4	12	-8	3.3	0.037
	appetitive conditioning	Amygdala	R		28	-2	-14	3.09	0.032
	[$CS_{+2€} > CS_{+1€} > CS_{0€}$]	VS	L		-4	12	-8	3.55	**0.017***
			R		14	6	-12	3.33	0.034
	aversive conditioning	-							
	[$CS_{-2€} > CS_{-1€} > CS_{0€}$]								
Group differences									
AUD > HC	Pavlovian conditioning	Amygdala	R		28	-4	-22	3.74	**0.004***
	[$CS_{-2€} > CS_{-1€} > CS_{0€} < CS_{+1€} < CS_{+2€}$]		L		-24	-8	-22	3.35	**0.014***
AUD > HC	appetitive conditioning	Amygdala	R		26	-6	-22	4.06	**0.001***
	[$CS_{+2€} > CS_{+1€} > CS_{0€}$]		L		-24	-8	-22	3.56	**0.007***
		Hippocampus	R		26	-10	-22	3.76	**0.004***
			L		-24	-10	-24	3.5	**0.011***

684 p_{FWE}: family-wise error-corrected at p<0.05 for bilateral anatomical region; *denotes significance after Bonferroni
685 correction for number of ROI comparisons; L: left hemisphere, R: right hemisphere; VS: ventral striatum, AUD:
686 alcohol use disorder; HC: healthy controls
687

Figure Legend 1

A Exemplary appetitive conditioning trial. In each trial, a CS (fractal image combined with one out of five pure tones) was presented either on the right or left side of the screen for 3 s. After a fixed 3-second-trace interval, the associated monetary US (or neutral outcome (0 Cent)) appeared on the opposite site for 3 seconds (100% reinforcement schedule). Trials were separated by a jittered ITI (exponentially distributed; range: 2-6s; mean=3s). The paradigm comprised 5 different conditions (two appetitive, two aversive, and one neutral condition). CS assignment to conditions was counterbalanced across participants. **B** CS-US contingency knowledge. Mean probability of choosing the higher-valued CS during post-conditioning forced-choice task did not significantly differ between patients with AUD and healthy controls (W=2515.5, p=.77; AUD: n=75, mean(SD) = 85.6(18.1); HC: n=69, mean(SD) = 87.8(16.7)). Only participants performing significantly over chance (teal color-coded participants; i.e. over 50% correct choices, as confirmed by a binomial test) were considered contingency aware (83% of AUD patients, 91% of healthy controls) and included in the final sample (participant characteristics, see Table 1; see Supplementary Table S2 for sample characteristics of aware vs. unaware participants).

Figure Legend 2

Stronger differential BOLD responses in bilateral amygdala during Pavlovian conditioning in AUD patients compared to control participants (amygdala right: Z=3.74, p_{FWE_ROI}=.012; amygdala left: Z=3.35, p_{FWE_ROI}=.041). Group differences were driven by both increased BOLD responses toward Pavlovian CSs in AUD patients compared to healthy controls (p_{FWE_ROI}<.001), as well as increased BOLD responses towards the neutral cue in healthy participants compared to AUD patients (p_{FWE_ROI}≤.012; see Supplementary Material).Visualization threshold of T-map at T≥3.

Figure Legend 3

BOLD responses in left amygdala during Pavlovian conditioning were positively associated with subsequent PIT behavior in AUD patients compared to control participants (amygdala left: Z=3.35, p_{FWE_ROI}=.048). Visualization threshold of T-map at T≥3.

Figure Legend 4

A Significant group difference between prospective relapsers, abstainers, and controls during Pavlovian conditioning in the right amygdala ($F_{2,416}$=8.58, $p_{FWE\ ROI}$=.033). Both prospective relapsers (Z=3.47, $p_{FWE\ ROI}$=.033) and abstainers (Z=3.40, $p_{FWE\ ROI}$=.042) showed increased BOLD responses compared to control participants, while patient groups did not significantly differ. **B** Within patients who relapsed, differential amygdala responses during Pavlovian learning were inversely related to relapse latency (Z=2.94, $p_{FWE\ ROI}$=.047). Visualization threshold of F-/T-map at F≥6/T≥3.